Task 8637
cont. # N00024-68-
C-1193
Purdue U.

(6) INTERACTIVE SONAR FEATURE EXTRACTION.

MOST Project -3

(1) B.S.

(9) Quarterly Progress Report.

by

(10)
E. A./Patrick,

L. Y. L./Shen,

F. P./Fischer

G./Carrayannopoulos

(11) 1968

(15) N00024-68-C-1193

(12) 36p.

Purdue Univ.
Lafayette

291 650

LB

CONTENTS

I. Introduction

A. The Unified Problem

The problem considered is where n samples of data are to be analyzed and classified. A particular sample of the data is, in general, two dimensional. For example, a sample may consist of d digitized time functions, each waveform the output from a hydrophone transducer. If the hydrophones are arranged along a straight line which is called the space dimension, then a sonar sample may be considered a discretized space-time waveform.

Our objective is to use the n samples to construct a mapping from the data observation space (the digitized data is assumed to consist of K_s data points where s indicates sample s) to what will be called the class space. There are M classes in the class space and M may be an unknown quantity. Mathematically, the M classes may be considered as points on the real line.

Historically, statistical learning theory and pattern recognition have produced certain operations from which the above mapping from the observation space to the real line is constructed. Before considering these operations, lets try to construct a single mapping from the raw data space to the class space. Suppose there are n samples and each consists of K data points. For example, let K=3 as shown in Fig. 1. In this example we further assume the n samples form 3 clusters

(M=3) in the observation space. If these three clusters could be identified and a mapping constructed which would classify any subsequent sample as belonging to one of these three classes, our problem is solved. Unfortunately this is not so easy, especially when K>>3.

One way to identify the clusters is to construct a mapping as follows: map point 1 in class ω_1 onto the real line at position 1 as shown in Fig. 2. Next let the height of the line at position 2 correspond to the distance between point 1 and its nearest neighbor, point 2. Continue this procedure and note that the nearest neighbor to point 9 in class ω_1 is point 10 in class ω_2; thus there is an abrupt "jump" in the amplitude plotted in Fig. 2. The jump at point 10 in Fig. 2 represents a separating boundary between class ω_1 and class ω_2.
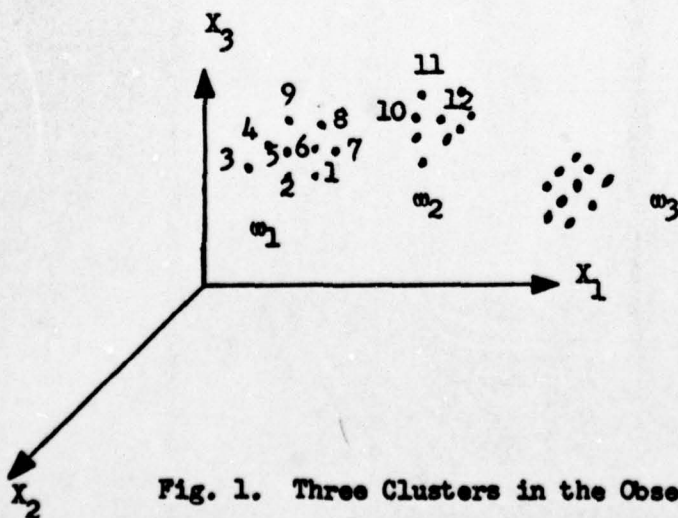

Fig. 1. Three Clusters in the Observation Space

Nearest Neighbor Distance

```
     1  2  3  4  5  6  7  8  9  10  11 12
```

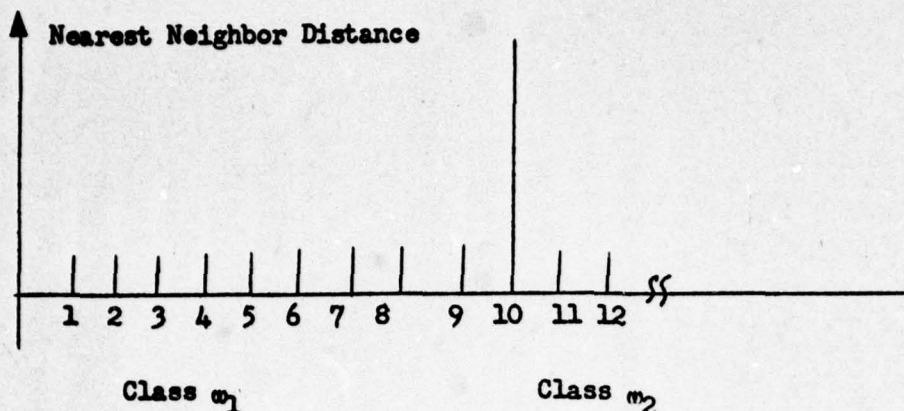Class ω_1 Class ω_2

Fig. 2. Nonlinear Cluster Mapping Onto Real Line

The above mapping is a clustering algorithm which makes it possible to identify clusters. For example, if Fig. 2 appeared as a computer output display, then the samples to the left of the jump can be identified as class ω_1; this is called <u>interactive identification</u>.

A general family of cluster maps from V^L to V^ℓ is described as follows: assume

(a) there is one distinct mode for each class, or the classes are "nearly" separable.

(b) The dimensionality is much greater than the number of samples. Let D_L^i be the distance in V^L between the ith pair of samples and D_ℓ^i the distance in V^ℓ after some transformation of the samples. Then we wish the transformation to be such that

$$\sum_{\text{All } i} \left(D_L^i - D_\ell^i\right)^2$$

is minimum. It may then be possible to display the classes as clusters[*] in ℓ dimensions, say on a computer output display screen if $\ell=2$. If there are two clusters, then we have identified two classes; we might, for example, replace each class by its mean vector and covariance matrix.

B. Dimensionality Reduction

If there is one or more relatively noisy dimensions compared with other dimensions, cluster algorithms as described in the last section may not work. Mathematically, the reason is that the distance criteria used by the clustering mapping must fit the data. Such a shaping of the distance criteria is fundamental to all estimation/recognition algorithms including the tolerance region approach or Kth nearest neighbor.

Shaping the distance metric is a general description of the operation of dimensionality reduction and feature selection.

There is another operation, different[**] from that of feature selection or simple dimensionality reduction, which involves the interpolation of a waveform or sequence of data

---

[*] There are cases, such as the case of a noisy dimension, where distance preserving, cluster mappings will not correctly display clusters unless distance is defined to include certain structure of the data.
[**] The difference is in terms of apriori knowledge.

points. This operation is called feature extraction.

C. Feature Extraction

The feature extraction operation requires apriori knowledge to get started. Because waveforms are continuous in time and space (when an array of transducers provides the signals), there are an uncountable number of sets of apriori knowledge that can be used for feature extraction. In order to reduce the number of sets and the size of the sets, certain "fuzzy" structure of the waveforms can be utilized; for example, partitions can be used to isolate modes of the waveforms. Then the waveform in each partition can be expanded with a "few," relatively simple basis functions. Approximating waveforms by partitioning and expansion is not a new approach; in fact there are mathematical functions called spline functions which follow this approach. Feature extraction for statistical waveforms is not, however, only a matter of approximation or waveform representation. Each waveform in a class of waveforms must be approximated. The method of partitioning and expansion must provide for sufficiently adequate representation of each waveform so that any waveform in the class can be distinguished from any waveform in another class. That is, partitioning and expansion should result in extracting intraclass properties from each class as well as extracting interclass properties, which provide for distinguishing between members of different classes.

Partitioning and expansion is a nonlinear mapping from an infinite (or high) dimensional, waveform observation space to a finite dimensional subspace. The nonlinear mapping should be such that waveforms from a class are mapped "close" together in the subspace while waveforms from different classes are mapped far apart. Reasons for a feature extraction mapping are discussed in the following subsection.

WHY USE A FEATURE EXTRACTION MAPPING

Theoretically, estimation and decision making can be performed in the original, high dimensional space. For example, a Kth nearest neighbor rule or elliptical tolerance region decision rule can be directly applied without feature extraction. However, there are good reasons for determining a feature extraction mapping which are summarized below.

1. A feature extraction mapping allows decision making to be accomplished in a lower dimensional space than the original waveform observation space; thus decision making can be extremely much faster using the feature extraction mapping.

2. Experimentation can take an unacceptably long amount of time without a feature extraction mapping. Since the objective is to evaluate the performance of an automatic classification system, it must be possible to process unclassified samples in a reasonable length of time.

3. Total system performance can be improved by feature
extraction (it also can be improved by feature selection).
This is a result of the fact that eventually the probability
density function of each class must be estimated; this estimation
can be deteriorated by noisy dimensions having little classification
information.

4. Before estimation and recognition can begin, an observation
space must be defined. Theoretically, this observation space
must be fixed. Lack of exact knowledge of synchronization gives
rise to a non-fixed observation space unless synch-free
features are extracted. In practice, a "fuzzy" observation
space is acceptable and provides a means of obtaining a fixed
observation space. Partitioning is one way to concentrate
attention on "local" properties of waveforms; then, in a given
partition, "fuzzy" properties of waveforms in that partition
can be extracted.

D. Clustering and Estimation/Recognition

Feature extraction is not a stand-alone operation. It
must be accomplished with the objective of following it with
an estimation and recognition operation. Eventually, we must
estimate or experimentally determine the shape of tolerance
regions used to measure the probability density for each class,
in the vicinity of an unclassified test sample (candidate
sample). The geometrical shape of the tolerance region is

partly determined by feature selection. Feature extraction
makes the feature selection operation manageable. After suc-
cessfully accomplishing feature extraction, feature selection,
and estimation/recognition, we can be concerned with the further
complexity reduction technique of storing only the decision
boundary in the observation space rather than the mapping and
all the mapped samples.

The operations of feature selection and estimation/recognition
can be more complicated when the data for any class is multimodal
than when it is unimodal. Therefore it can be very helpful
and instructive to locate modes or clusters in data. A clustering
algorithm can be used to map multidimensional data onto a two
dimensional computer output display (Calcomp plotter or storage
screen). Such clustering mapping is not in general feature
selection. Rather, it is used to identify modes. After
identifying a mode, the samples constituting the mode can be
replaced by a mean vector and covariance matrix (corresponding
to the unimodal assumption). Also, significant directions
between modes can be found, which is a simple form of feature
selection. The vector corresponding to such a significant
direction is then a dimensionality reducing filter.

E. Networking

Even with successful feature extraction, the dimensionality
of the data may be too high for feature selection or estimation

and recognition to be simply accomplished. In this case, subsets of the features are processed individually. The decision made using one particular subset is then used as an input feature to be used in another subset. A decision is then made using the latter subset and the decision from the previous subset. More complicated methods of networking can be used.

In the following section, an example of feature extraction based on partitioning and expansion technique is described. Because the method of partitioning can be determined by an operator observing sonar waveforms on a computer output display, it is called interactive feature extraction.

II. Automatic Feature Extraction From Two Dimensional (Space Time) Data

There has been considerable recent interest in interpolation functions for irregularly-spaced data [1,2,3,4]. Work has been primarily concerned with displaying data in some type of contour map with the objective of being able to compare the display with a display obtained using other data. Such work has emphasized the use of two-dimensional interpolation functions which smooth the data displayed. The literature cites such reasons for display as to analyze the data for extremes, gradients, etc.

This section is concerned with the problem of mapping two dimensional, supervised data sets to one dimension for subsequent classification of candidate data. There are M classes, a two dimensional data sample consists of K data points, and each

class contains N samples. The jth data point, common to all M classes, is characterized by the coordinates $D_j = (x_j, y_j)$. If the value of the jth data point for the ith class is z^i_j, then the triplets (x_j, y_j, z^i_j), $j = 1, 2, \ldots K$ describe one sample from the ith class. Since the ith class contains N samples, we define $\left\{ \left\{ (x_j, y_j, z^{i\,s}_j) \right\}^K_{j=1} \right\}^N_{s=1}$ as the N sets of data points (N samples) for the ith class.

AUTOMATIC PARTITIONING

Let D_e be the first data point examined, as seen in Fig. 3. Select the g nearest data points to D_e. Using these g data points and an apriori set of basis functions $\left\{ \varphi_0(D_j), \varphi_1(D_j), \varphi_2(D_j), \varphi_3(D_j) \right\}$, calculate Fourier series coefficients for the ith class for one sample:

$$a^i_o = \sum_{j=e-\frac{g}{2}}^{e+\frac{g}{2}} z^i_j \, \varphi_o(D_j)$$

$$a^i_1 = \sum_{j=e-\frac{g}{2}}^{e+\frac{g}{2}} z^i_j \, \varphi_1(D_j)$$

$$a^i_2 = \sum_{j=e-\frac{g}{2}}^{e+\frac{g}{2}} z^i_j \, \varphi_2(D_j)$$

$$a^i_3 = \sum_{j=e-\frac{g}{2}}^{j=e+\frac{g}{2}} z^i_j \, \varphi_3(D_j)$$

Form the vector

$$\underline{a}^i = [a_0^i, a_1^i, a_2^i, a_3^i]$$

which is <u>for the specific data point D_e using g neighboring</u>
data points to D_e. Since there are N_i samples for the ith
class, the preceeding can be accomplished for all N samples
to obtain

$$a^{i_s}, \ s = 1,2,\ldots N._i$$

An example using the spanning set $1,x,x^2,x^3$ is shown in
Fig. 4. In this example a simple feature selection technique
is utilized to map the four dimensional vector space containing
$\underline{a}^i$ to the one dimensional vector between the class means.
Using a definition of similarity supplied by the operator, it
is determined if D_e is similar, with respect to the features
of each class associated with it, to D_{e-1}. If it is, D_e and
D_{e-1} are said to be in the same partition; otherwise a new
partition is created.

After all K data points are examined, the output is the
partition locations and the corresponding $\overline{a}^i$, i= 1,2, for the
partitions. Thus feature extraction has taken place with the
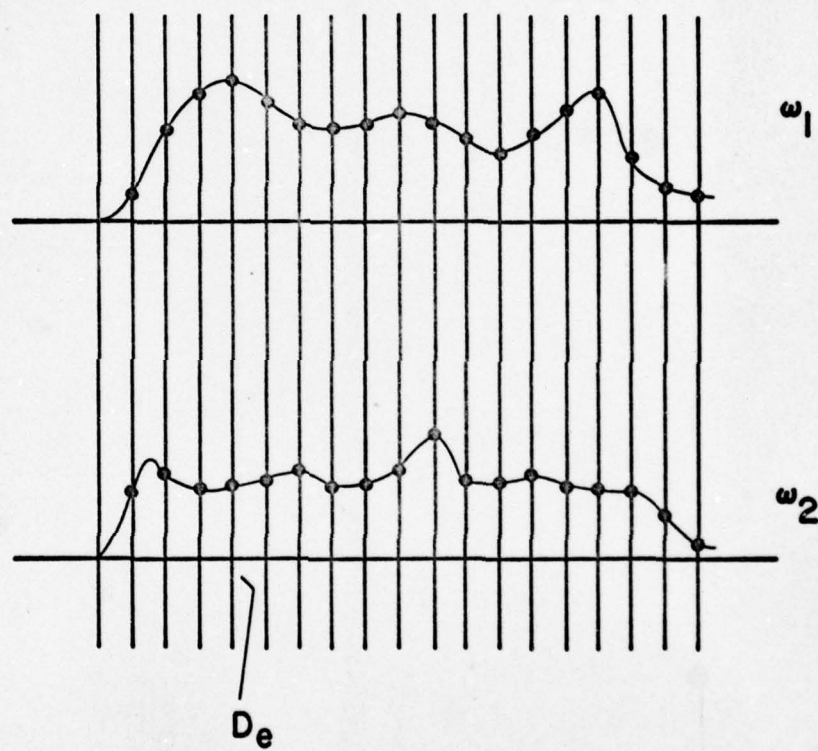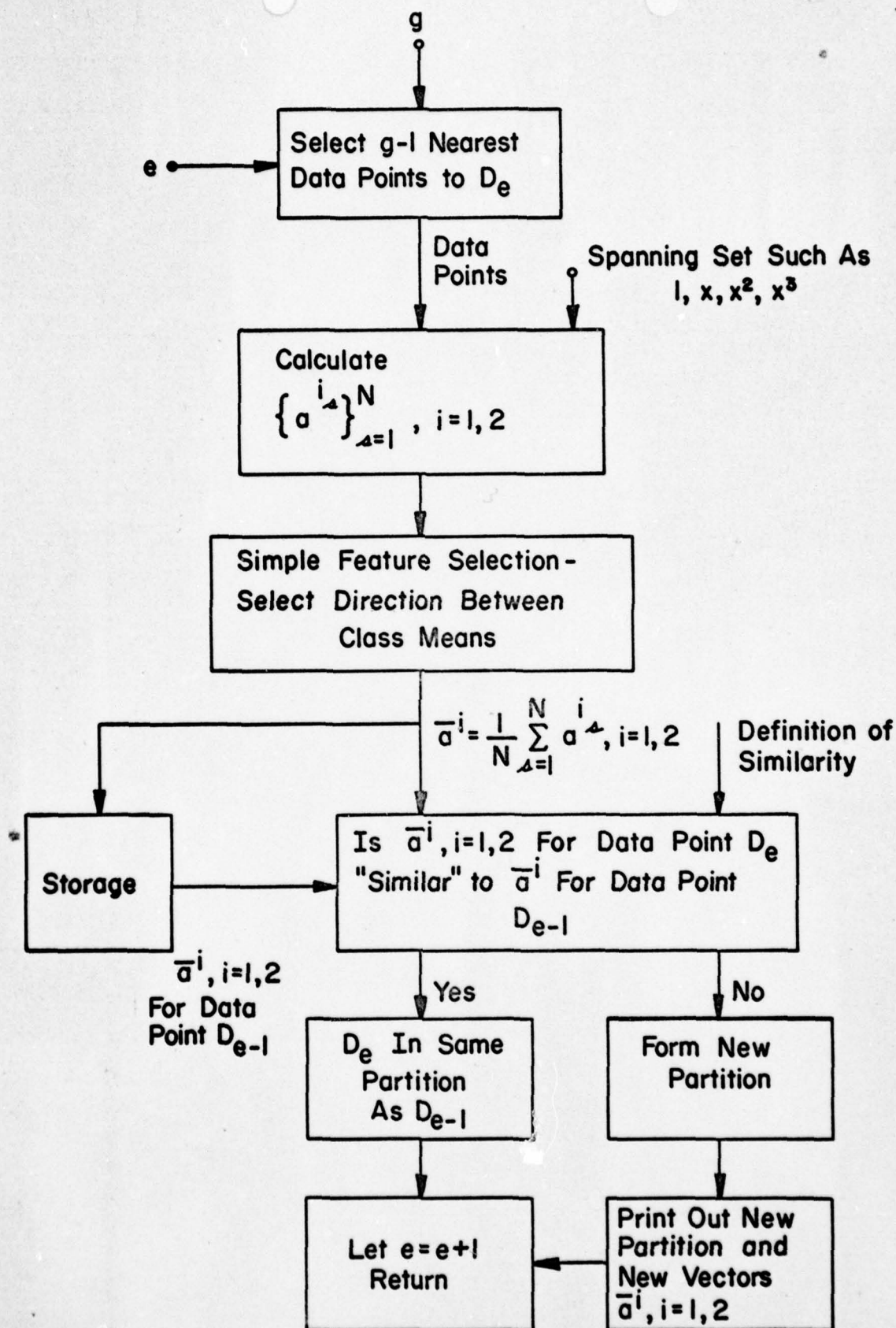operator supplying g, the spanning set, and the definition
of similarity.

FIGURE 3.

g

Select g-1 Nearest
Data Points to D_e

e

Data
Points

Spanning Set Such As
$1, x, x^2, x^3$

Calculate
$\left\{ a^{i}_{\Delta} \right\}^{N}_{\Delta=1}$, $i=1, 2$

Simple Feature Selection -
Select Direction Between
Class Means

$\overline{a}^i = \dfrac{1}{N} \sum\limits_{\Delta=1}^{N} a^{i}_{\Delta}$, $i=1,2$

Definition of
Similarity

Storage

Is $\overline{a}^i$, $i=1,2$ For Data Point D_e
"Similar" to $\overline{a}^i$ For Data Point
D_{e-1}

$\overline{a}^i$, $i=1,2$
For Data
Point D_{e-1}

Yes

No

D_e In Same
Partition
As D_{e-1}

Form New
Partition

Let e=e+1
Return

Print Out New
Partition and
New Vectors
$\overline{a}^i$, $i=1,2$

FIGURE 4. AUTOMATIC FEATURE EXTRACTION.

III. Experimental Results

The automatic feature extraction technique described in the previous section has been implemented as a Fortran IV softwave program. The program is written for operation in a CDC 6500 Computer with CDC 252 Computer Output, light pen, display screen.

In an initial experiment, five waveforms from each of two classes were drawn by the experimenter using a light pen. These waveforms simulated five sonar echoes from subs and five sonar echoes from non-subs. The algorithm automatically placed two partitions along the time axis. The basic functions utilized in each partition are Lagender Polynomials.

One of the waveforms from the simulated S class is shown in Fig. 6 and one of the waveforms from the simulated N class is shown in Fig. 8 Without using partitioning, a convential method of signal representation gave the approximations shown in Fig. 5 and Fig. 7 (for a fixed number of coefficients). Using the automatic partitioning technique, the approximation for the waveforms concerned is as shown in Fig. 6 and Fig. 8

It should be emphasized that the new technique accomplishes feature extraction using both interclass and intraclass properties of the data.

Next we give an example of clustering using computer graphics. Three classes of data were generated. The data samples were two dimensional and appeared as shown in Fig. 10. The actual

class labels of the data was as shown in Fig. 9, with A's, B's, and C's being the respective data sample labels. The computer did not know the data sample labels; ie the three classes were unsupervised. The objective of the cluster algorithm was to determine how many classes there are and separate them. After one application of a cluster algorithm, the data appeared as in Fig. 11. After seven applications of the algorithm, the data appears as in Fig. 12. Thus the three classes were uniquely separated. Such cluster algorithms are useful when used in conjunction with such algorithms as the automatic feature extraction algorithm.

IV. Target Signatures and Array Processing

We have previously discussed (Final Report on Contract N00024-67-C-1162) the spacial (array) steady state response of a target using the method of geometrical acoustics. A more complete description would be the space/time impulse response, which is discussed in this section. Space/time impulse response studies are helpful in determining how to apply automatic space/time feature extraction.

REVIEW OF PREVIOUS WORK

This section will be primarily devoted to a discussion of the previous report by Purdue [TR-EE68-21] on the geometrical acoustics approach to scattering. The work contained in that report is based on the method of geometrical acoustics for

calculating the acoustic pressure field returning from a target
to the elements of a receiving array. This is a simple and
useful method as long as the wavelength of the signal is small
compared to the radius of curvature of the target. The results
of the report point out the possibility of determining target
bearing and shape* from the received signal along a linear array.
The Spatial Nyquist rate and spatial Fourier transform which
was used in that report, are interesting ways of expressing well
known results from antenna array theory and it may prove useful
to mention these alternate interpretations.

The operation of taking the spatial Fourier transform
of the pressure distribution along the array is simular to adding
up the outputs of the individual elements of the array and
plotting this sum as the beam is scanned, with respect to the
line of the array, by shifting the phase of each of the elements.
An important result from array theory is applicable here, namely
the use of an amplitude taper across the aperture of the array
to reduce the sidelobe level of the beam thus formed. This
would be of particular importance in a reverberation environment
to improve the signal-to-noise ratio by reducing the response
due to reverberation from unwanted directions. This amplitude
taper would also be of benefit when attempting to resolve closely

_____

*No particular attention was given to the target aspect angle and length.

spaced and multiple targets, since there would be less clutter due to the sidelobes of adjacent targets.

The use of a spatial Nyquist rate for locating elements along the array corresponds to spacing of array elements to avoid bringing additional main lobes of the array factor into the visible range. This spacing can be larger than the half-wavelength distance, but caution is then necessary to avoid scanning an additional main lobe of the array factor into the visible range when trying to locate targets by taking a spatial Fourier transform.

There are a few areas where some caution is required in interpreting and extending the results of that report. The first of these is the fact that geometrical acoustics is not suitable for predicting the target response for the range of wavelengths comparible to the size of the target. It is in this range of wavelengths that some of the features of a target's response are to be found. This can easily be seen by examining the approximate impulse responses computed by the method of Bennett and Weeks which are presented in the next section. A second point that should be noted is that the results obtained for locating targets by taking the spatial Fourier transform are very sensitive to the phase of the signal at each of the hydrophones[*]. In many cases the phase front arriving at an array might be greatly distorted by variations in the media

---

[*] Therefore analysis should provide for noise added to the phase at each hydrophone.

across the aperture of the array.

RECOMMENDATIONS BASED ON TARGET IMPULSE RESPONSE CALCULATIONS:

To make use of impulse response calculations for perfectly
conducting electromagnetic scatterers, the acoustic wave equation
and boundry conditions were compared to the wave equation for
electromagnetic waves and a correspondence was established for
the case of two dimensional (cylindrical) scatterers. Transverse
Electric (TE) scattering from a conductor is simular to acoustic
scattering from a hard boundry. Transverse Magnetic scattering
(TM) is simular to acoustic scattering from a soft boundry.
The three dimensional case for acoustic scattering could not
be handled readily by the method available for electromagnetic
scatterers. However, the results obtained from the two dimensional
case should be adequate for reaching some basic conclusions.
These can be compared with classical results for a sphere to
relate them to three dimensional cases.

The following illustrations should be consulted at this
time: figures 13, 14, 15, 16, 17, 18, and 19. These figures
show the approximate impulse response in several directions.
The units along the abscissa of each response are in "light-
meters". For the electromagnetic case this is the time it takes
light to move one meter. For the acoustic case we could call
the units along this axis "speed-of-sound-meters". The impulse
responses shown are in the far-field region and the large

semicircle indicates the time when an impulse reflected from the origin of the coordinate system would arrive at the far-field region.

The reader will note that in Fig. 13 (circular cylinder of one meter radius, TE case) that the initial peak of the impulse response in the backscatter direction corresponds to a specular reflection from the front edge and could be predicted by geometrical optics or acoustics. The negative swing and subsequent positive swing could not be predicted by this method. It is interesting to note that this subsequent positive swing in the impulse response corresponds to a wave traveling around the backside of the cylinder and returning towards the source. This is known as a creeping wave. Simular interpretations can be found for this wave in the other TE (hard boundry) cases. For the TM cases (soft boundry) a creeping wave does not occur.

The frequency response of the hard circular cylinder (TE) case is shown in Fig. 20. Note that the response drops off quite rapidly for k (actually k a+1, since a=1) less than one. We note that k=w/v, or k=2π/λ where w is the radian frequency, v is the velocity and λ is the wavelength.

To carry these results over to three dimensions consult Fig. 21 which shows the ratio of acoustic to geometrical cross section of a fixed rigid sphere of radius a. It should be noted at this time that in actual practice with acoustic targets, ideal hard and soft boundries are not available. Additional

complications of flexure waves in shells and other resonances inside the object must be considered. These will result in variations in the calculated target response for that shape object. Another source of difficulty can be seen in Fig. 19 which shows the TE (hard boundry) response of a corner reflector. Note that the impulse response varies significantly in various directions. In a practical situation fins and concave surfaces on a scatterer would give similar behavior.

In spite of the limitations and difficulties indicated in the previous paragraph, the basic behavior of the impulse and frequency responses of targets gives an indication of several general features that can be exploited for target identification. For example, the target return from small scatterers such as fish bladders could be reduced by selecting frequencies such that the corresponding wavelength is large compared to the size of these targets so as to place them in the "Rayleigh Region" of the response curve (see Fig. 21). Since the targets of interest such as submarines are much larger, it should be possible to exploit this variation in target behavior as a feature for distinguishing these two types of targets. It might also be possible to determine something about target shape from a knowledge of the predicted impulse responses.

Since the impulse response of a target varies significantly in several directions, additional information for target

classification in a volume of interest could be obtained by investigating the response in several directions. The impulse responses of simple shapes such as cylinders, strips, and the sphere do not vary significantly in a small range of angles. However, concave and other more complicated shapes do exhibit considerable variation as a function of angle and this might be exploited as a feature for target identification.

To give an idea of the frequency ranges of interest for utilizing the above recommendations, we will consider cylinders of one meter radius, and ten meters in radius. Since the speed of sound is about 1500 meters per second in water, for a one meter radius cylinder a ka value of one would correspond to a frequency of 1500 radians per second or about 240 Hertz[*]. For a ten meter radius target a ka of one would correspond to a w of 150 radians per second or about 24 Hertz.

Figures 13 through 21 are from "A Technique for Computing the Approximate Electromagnetic Impulse Response of Conducting Bodies" by C. L. Bennett and W. L. Weeks. June, 1968, Purdue Technical Report, TR-EE68-11, NSF GK 2367.

---

[*] This corresponds to the break frequency in Fig. 20.

References

[1] Donald Shepard, "A Two-dimensional Interpolation Function for Irregularly-spaced Data," Proceedings of the 23rd National Conference Association for Computing Machinery, 1968, pp. 517-524.

[2] P. Switzer, C. M. Mohr, and R. E. Heitman, "Statistical Analysis of Ocean Terrain and Contour Plotting Procedures," Project Trident Report 1440464, Arthur D. Little, Inc., April, 1964.

[3] I. S. Benezin and N. P. Zhidkov, Computing Methods, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1965, Vol. I, Chapt. 2.

[4] C. de Boor, "Bicubic Spline Interpolation," J. Math. and Phys. 41, pp. 212-218, 1962.

[5] E. A. Patrick and F. P. Fischer, "Nonparametric Feature Selection," TR-EE68-24, Purdue University, August, 1968.
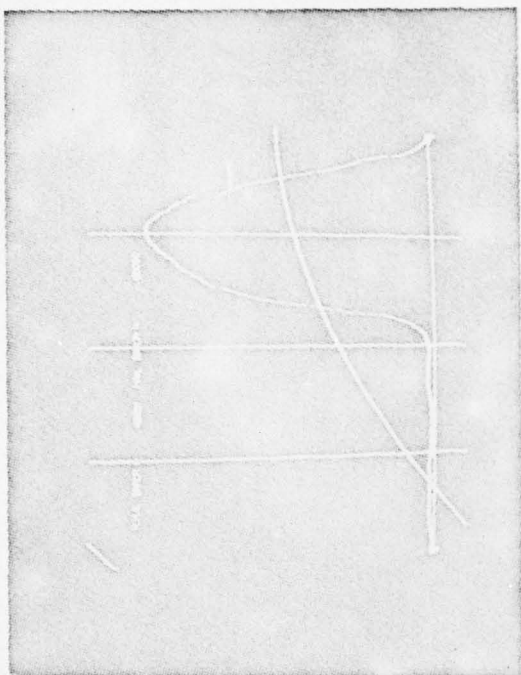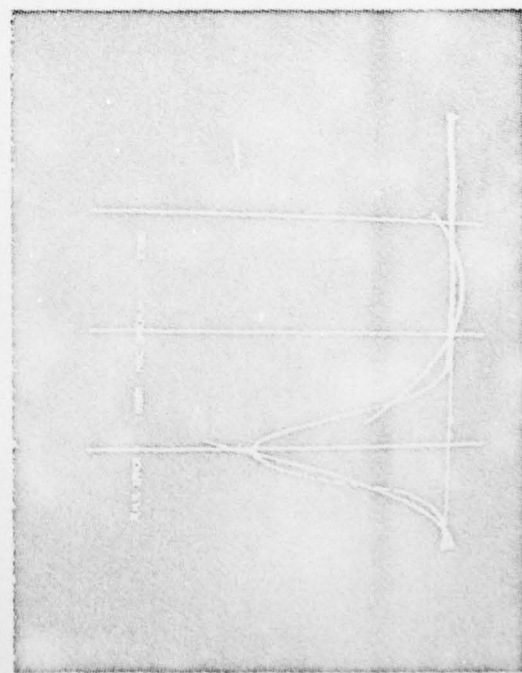
Figure 7

Figure 8

Figure 5
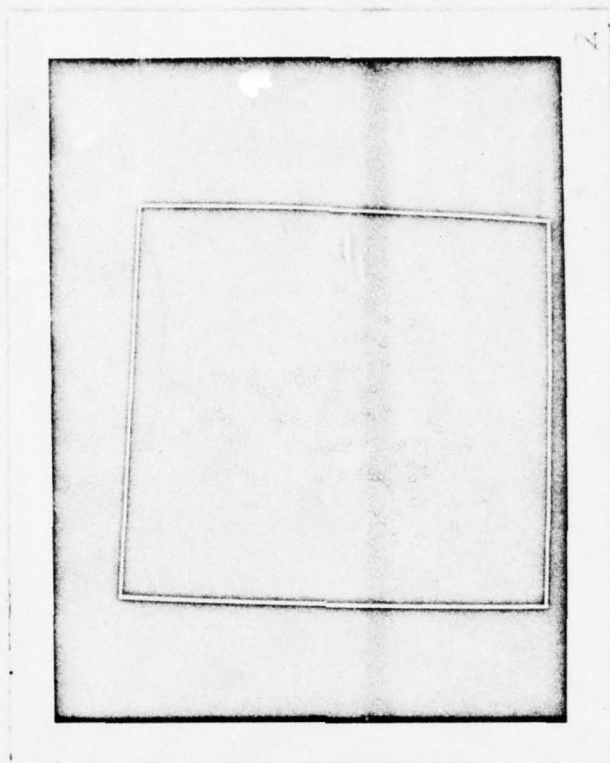
Figure 6
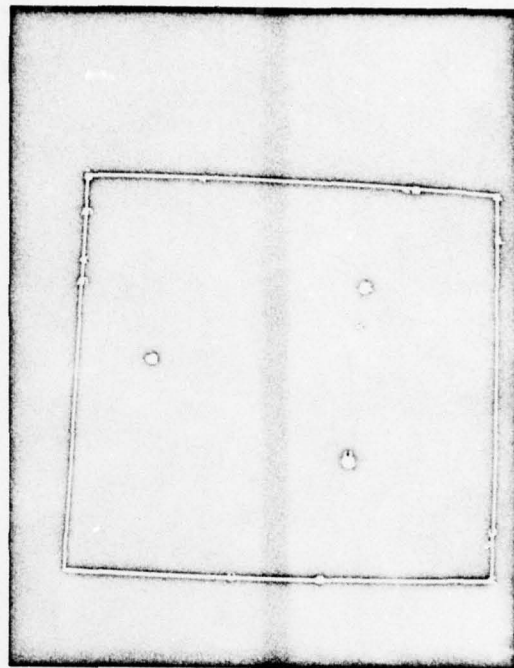
Figure 10
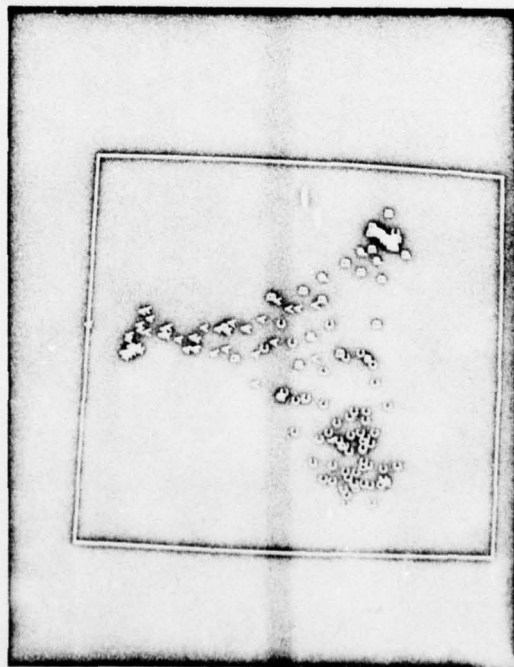
Figure 9

Figure 12

Figure 11
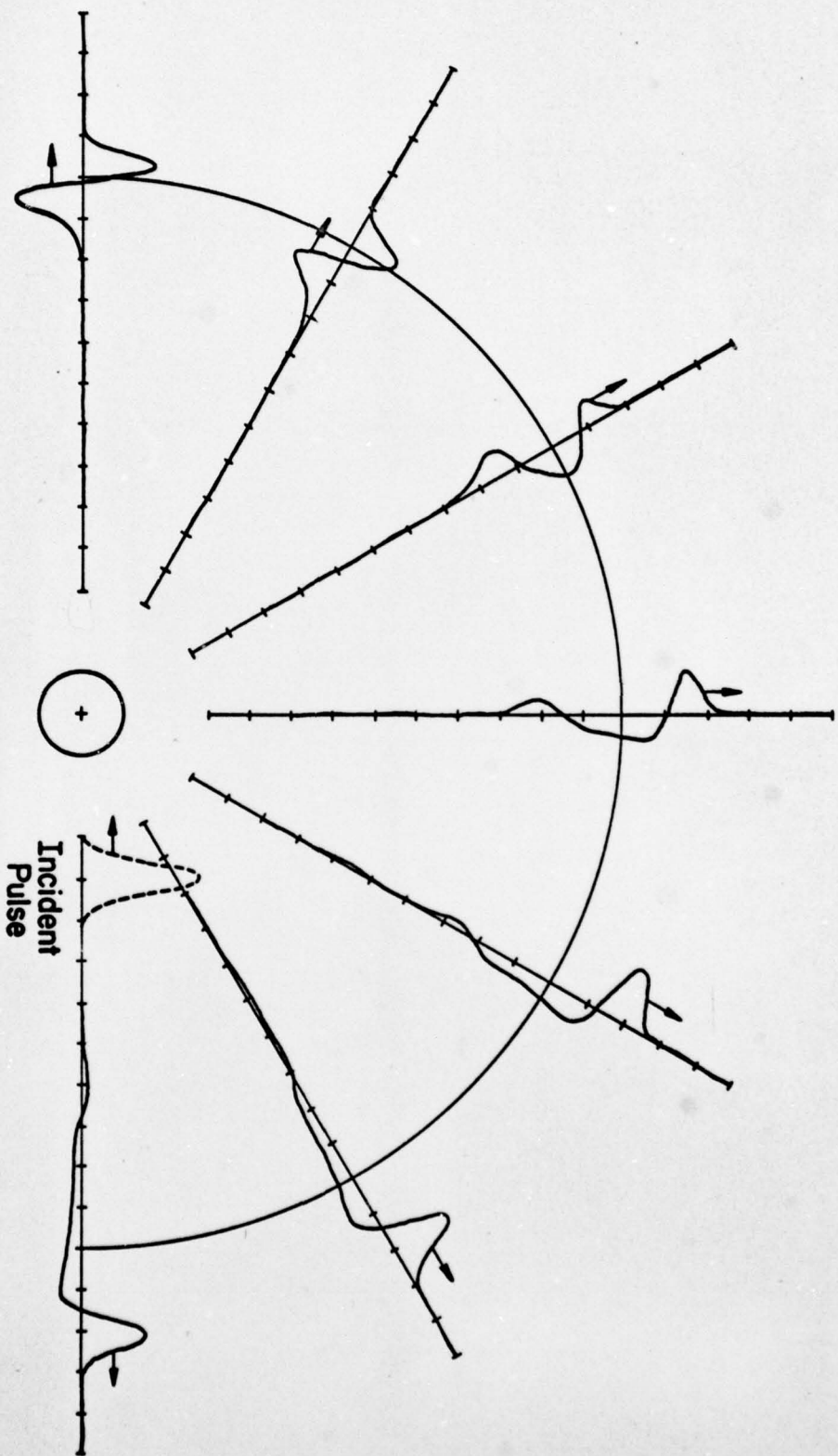
FIGURE 13. TE APPROXIMATE IMPULSE RESPONSE OF CIRCULAR CYLINDER.

Incident Pulse
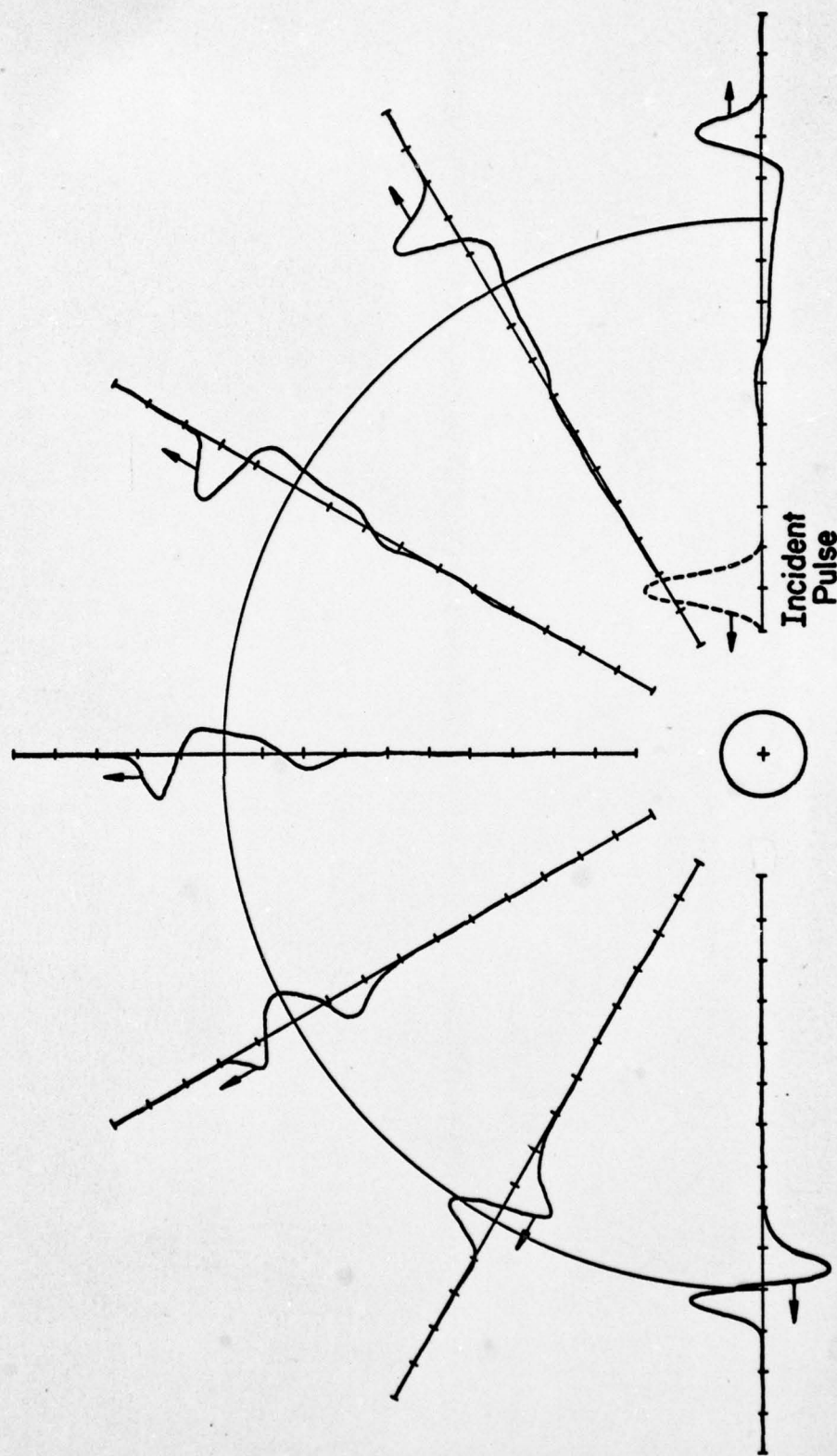
FIGURE 13. TE APPROXIMATE IMPULSE RESPONSE OF CIRCULAR CYLINDER.
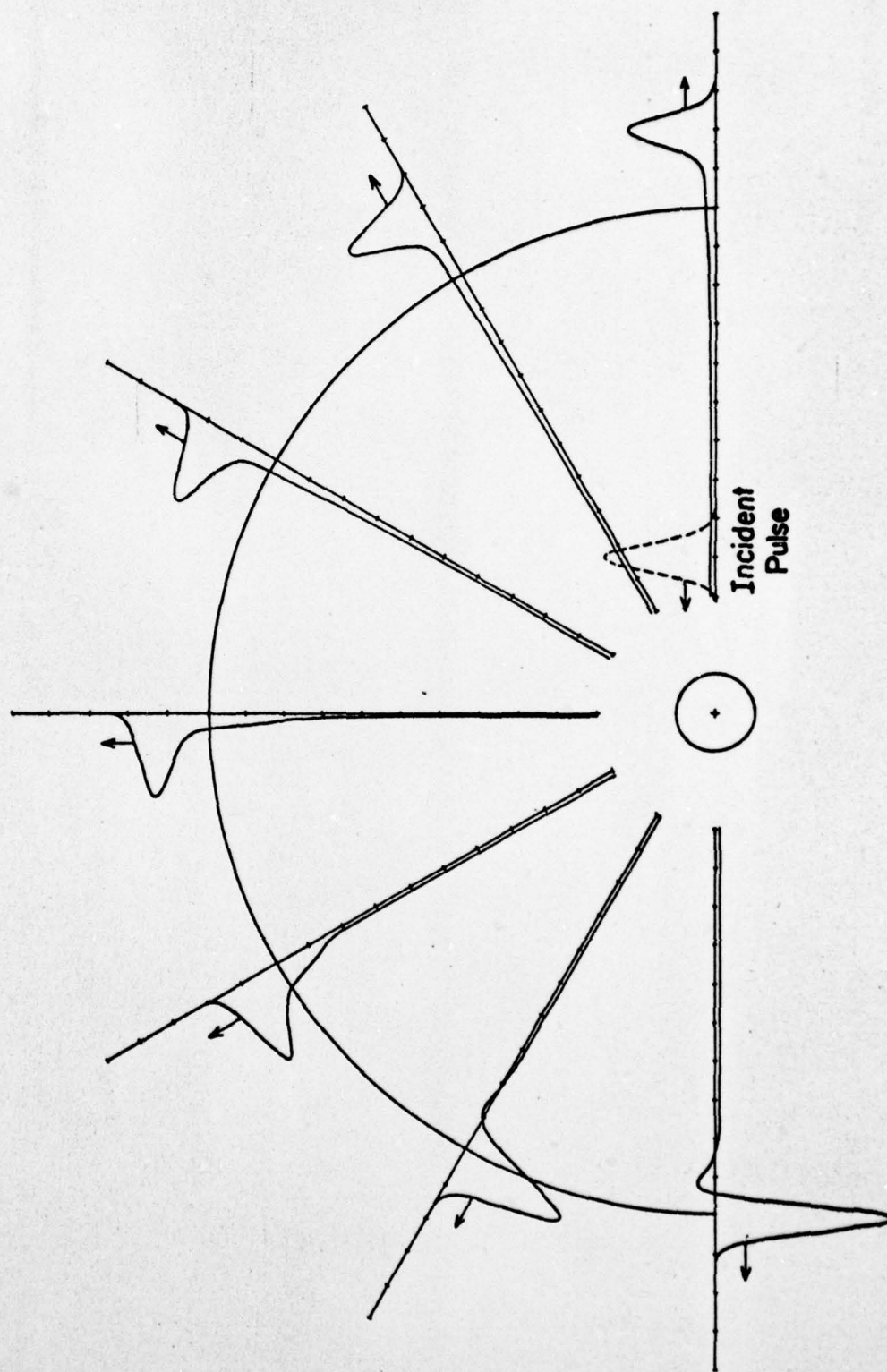
Incident
Pulse

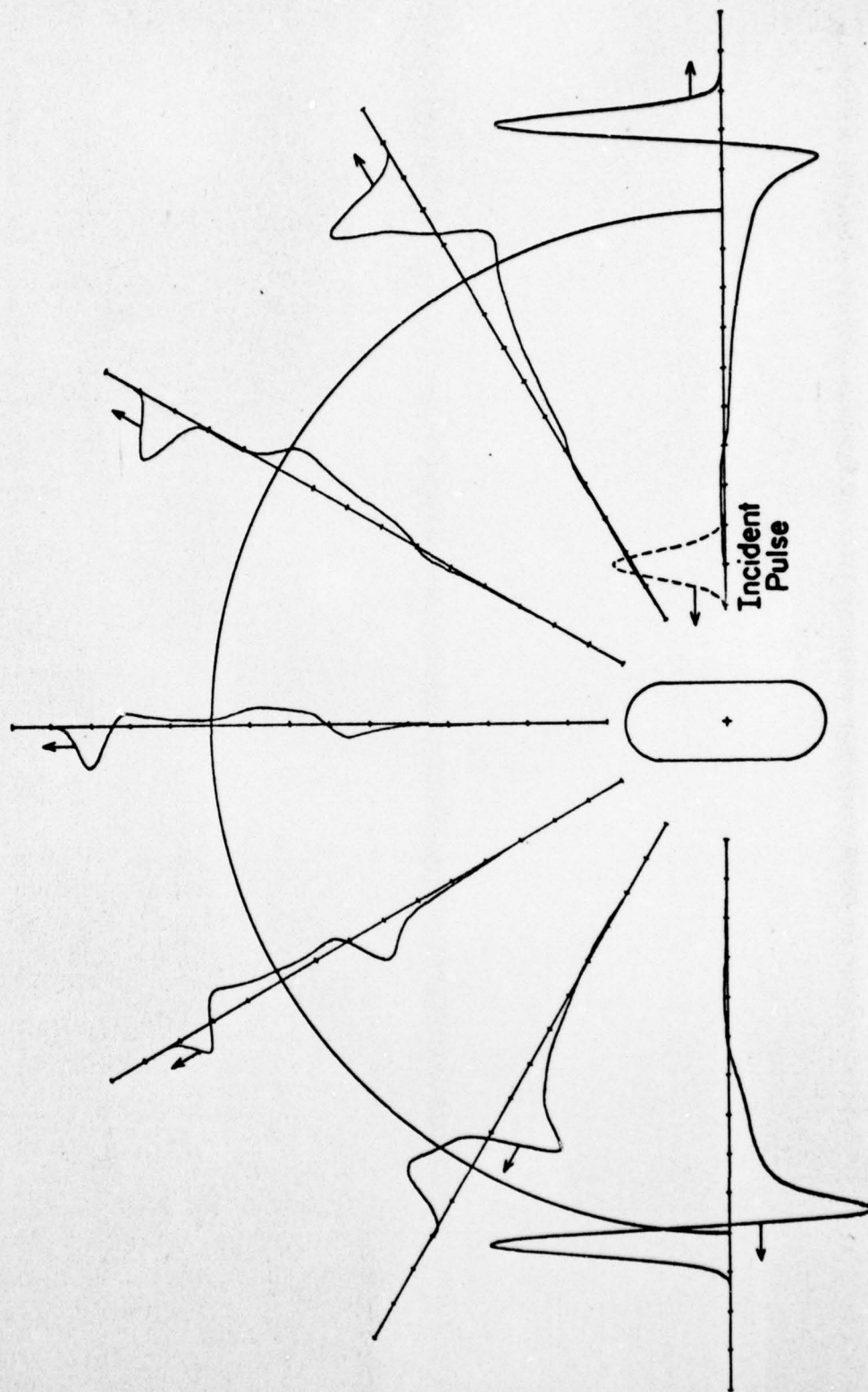FIGURE 14. TM APPROXIMATE IMPULSE RESPONSE OF CIRCULAR CYLINDER.

Incident
Pulse

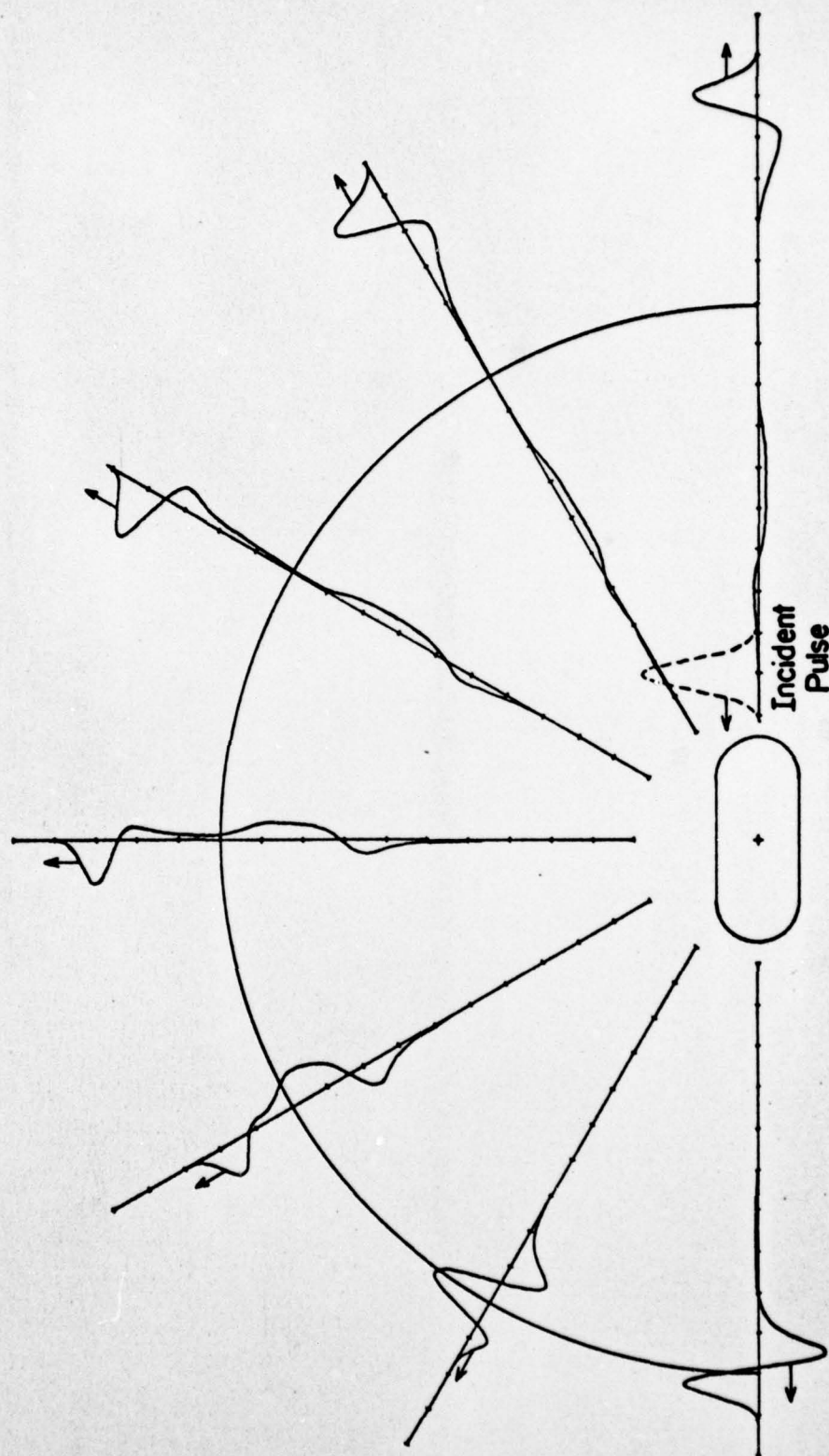FIGURE 15. TE APPROXIMATE IMPULSE RESPONSE OF STRIP WITH
BROADSIDE INCIDENCE.

Incident
Pulse

FIGURE 16. TE APPROXIMATE IMPULSE RESPONSE OF STRIP WITH END-ON
INCIDENCE.
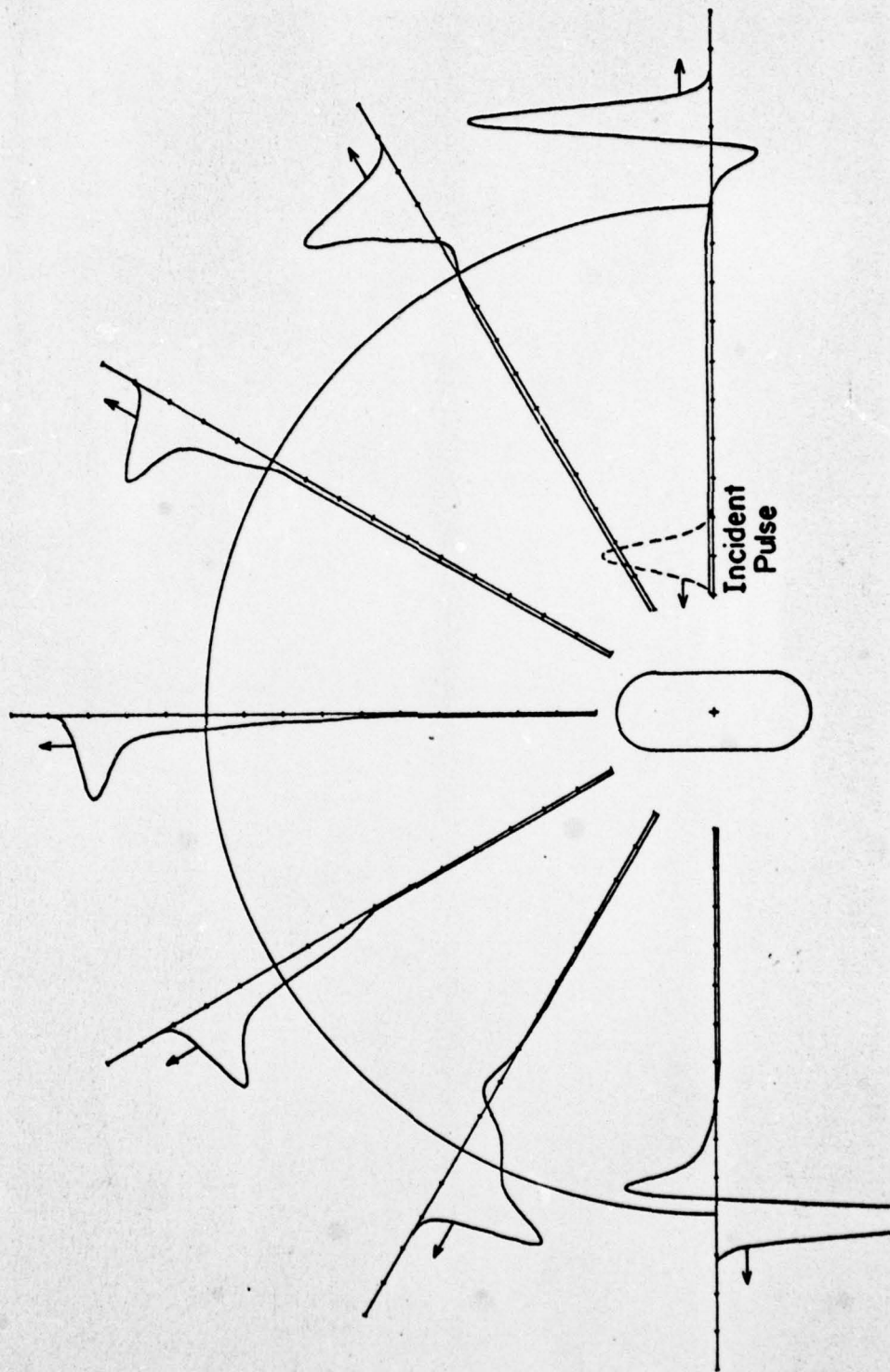
Incident
Pulse

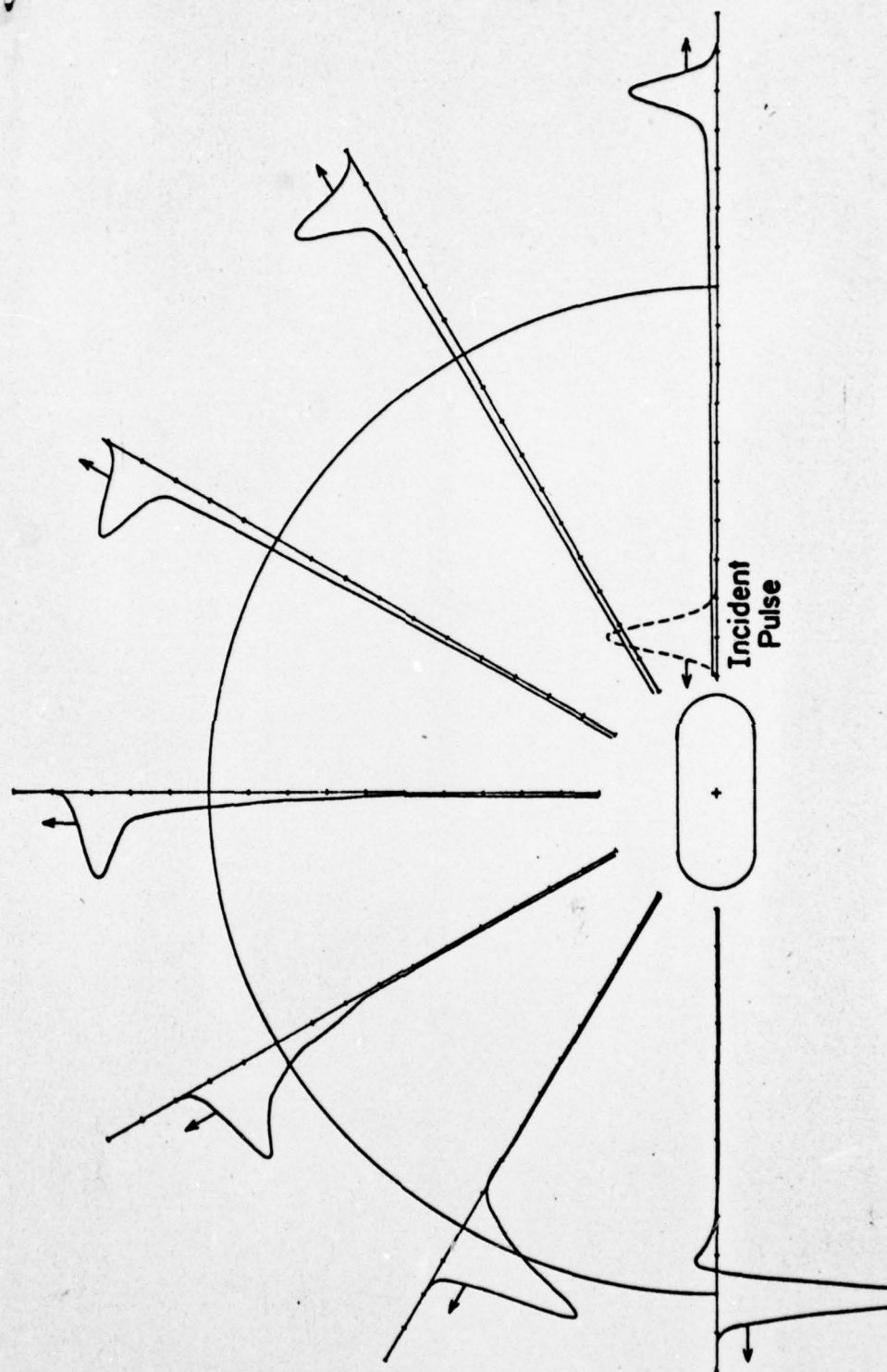FIGURE 17. TM APPROXIMATE IMPULSE RESPONSE OF STRIP WITH BROADSIDE
INCIDENCE.

Incident
Pulse

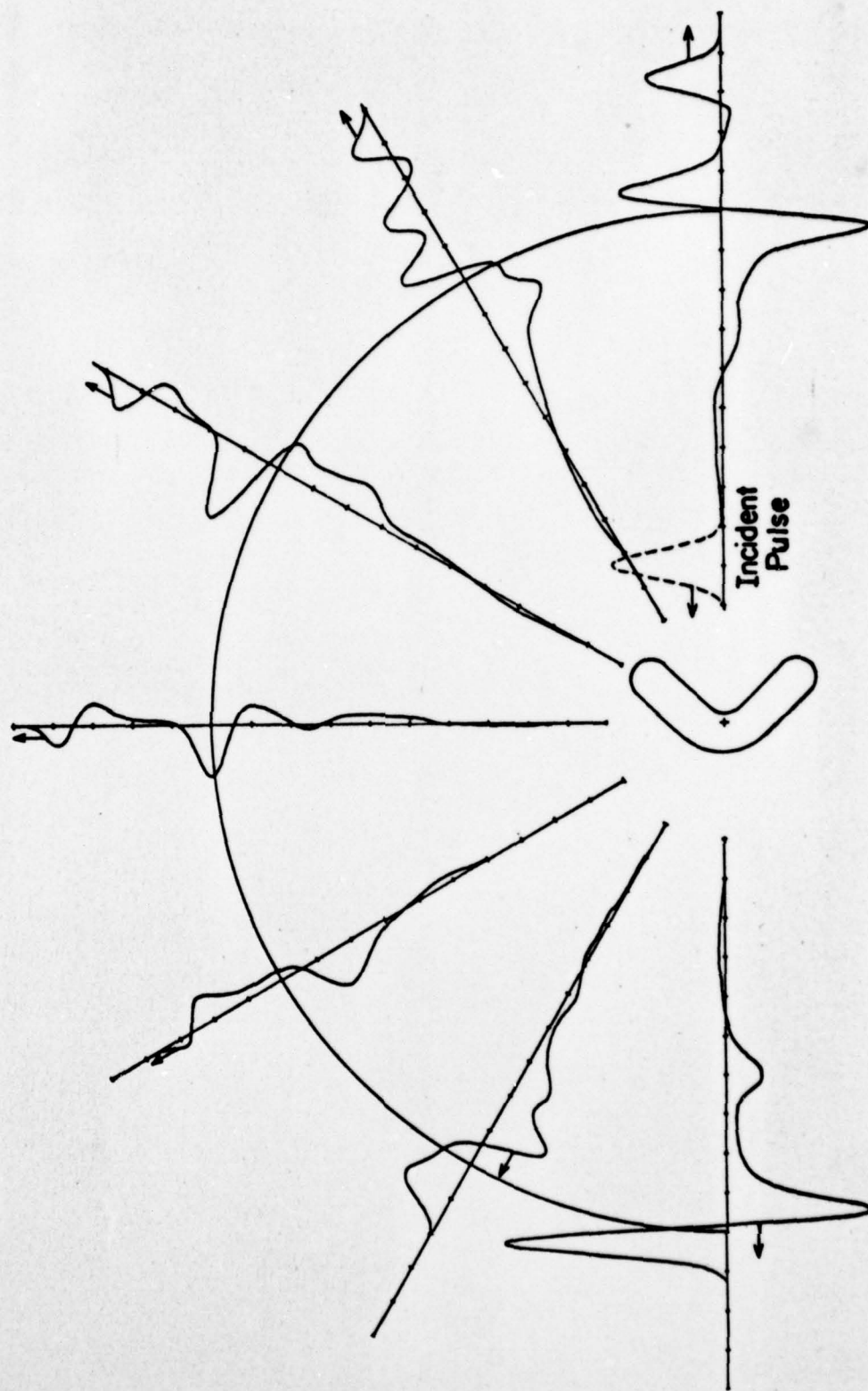FIGURE 18. TM APPROXIMATE IMPULSE RESPONSE OF STRIP WITH END-ON
INCIDENCE.

Incident
Pulse

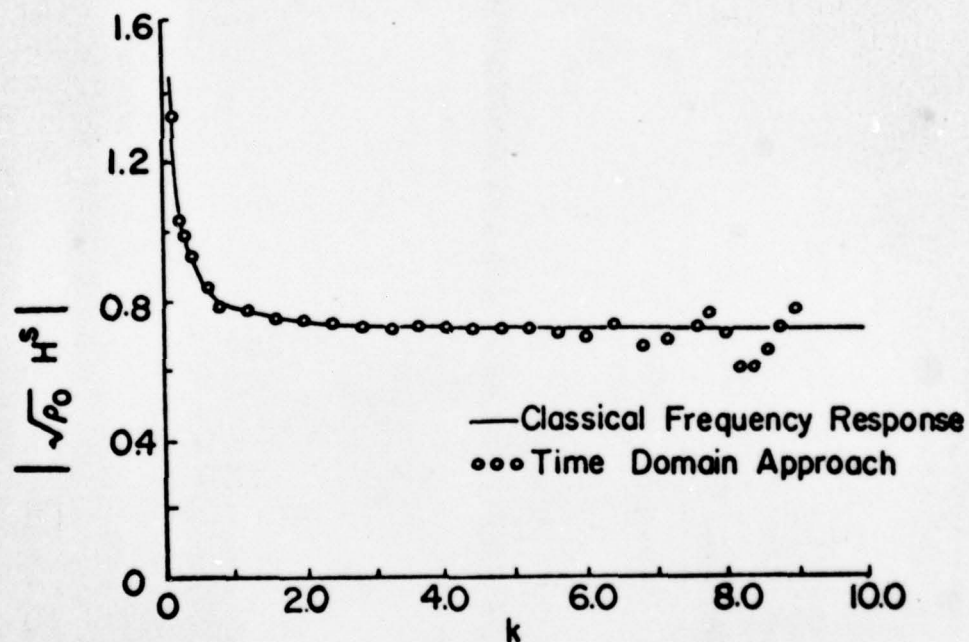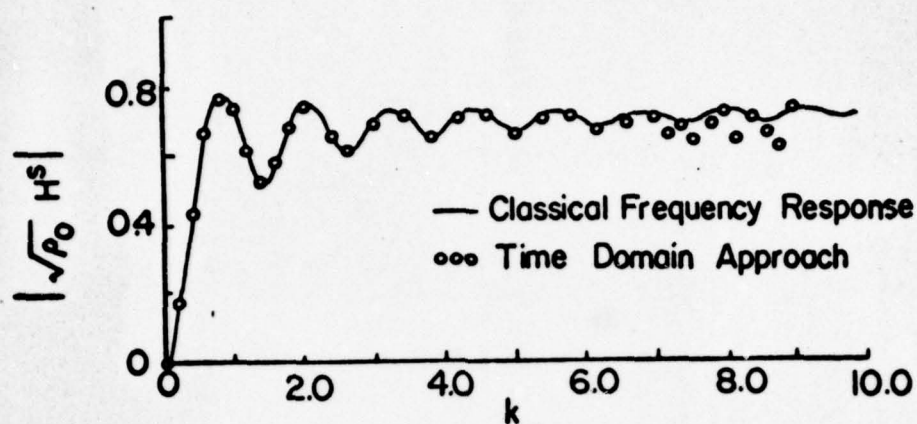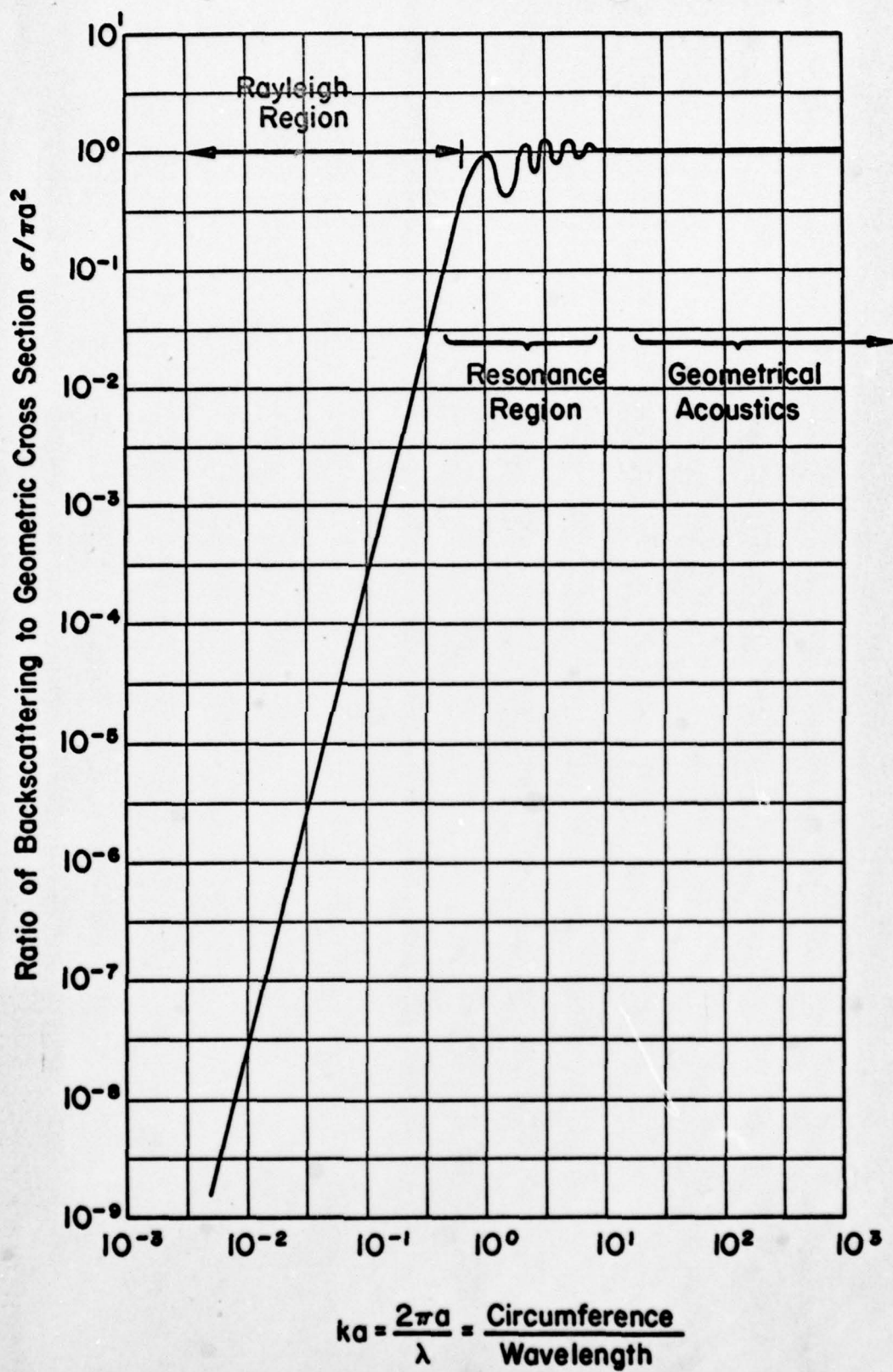FIGURE 19. TE APPROXIMATE IMPULSE RESPONSE OF CORNER REFLECTOR WITH FRONTSIDE INCIDENCE.

TE FREQUENCY RESPONSE IN BACKSCATTER
DIRECTION OF CIRCULAR CYLINDER WITH
ONE METER RADIUS

TM FREQUENCY RESPONSE IN BACKSCATTER
DIRECTION OF CIRCULAR CYLINDER WITH
ONE METER RADIUS

FIGURE 20.

FIGURE 21.